Acta Psychologica 138 (2011) 322–328

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/ locate/actpsy

On letter frequency effects

Boris New [a,*], Jonathan Grainger [b]

[a] Paris Descartes University, Sorbonne Paris Cité & CNRS, France
[b] CNRS & Aix-Marseille University, France

ARTICLE INFO

Article history:
Received 22 November 2010
Received in revised form 30 June 2011
Accepted 1 July 2011
Available online 19 August 2011

PsycINFO classification:
2300
2340
2820

Keywords:
Letter-frequency effects
Token frequency
Letter and word processing

ABSTRACT

In four experiments we examined whether the frequency of occurrence of letters affects performance in the alphabetic decision task (speeded letter vs. pseudo-letter classification). Experiments 1A and 1B tested isolated letters and pseudo-letters presented at fixation, and Experiments 2A and 2B tested the same stimuli inserted at the 1st, 3rd, or 5th position in a string of Xs. Significant negative correlations between letter frequency and response times to letter targets were found in all experiments. The correlations were found to be stronger for token frequency counts compared with both type frequency and frequency rank, stronger for frequency counts based on a book corpus compared with film subtitles, and stronger for measures counting occurrences as the first letter of words compared with inner letters and final letters. Correlations for letters presented in strings of Xs were found to depend on letter case and position-in-string. The results are in favor of models of word recognition that implement case-specific and position-specific letter representations.

There is now a general consensus in favor of letter-based accounts of reading words in languages that use alphabetic scripts. The key idea is that visual word recognition initially involves processing the identity and positions of the word's constituent letters (see Grainger, 2008, for review). In line with this general approach, models of reading typically postulate a letter level of representation separate from a word level of representation. For instance, the interactive activation model (McClelland & Rumelhart, 1981) postulates that when a word is presented it activates feature units, letters units and word units, hierarchically arranged in that order. This model explains word frequency effects by variations in the resting level activations of word units. The more often a word has been seen, the higher its resting level activation. This explains why high frequency words are easier to recognize than low frequency words.

McClelland and Rumelhart (1981) chose not to vary the resting level activation of letter representations as a function of their frequency of occurrence. However, letters, just like words, vary considerably in their frequency of occurrence in print. Vowels occupy the top of the frequency hierarchy, but there is also a large variation in consonant frequency as can be seen in Fig. 1. Hence, to the extent that word frequency affects word recognition, one might expect letter frequency to affect letter recognition. More precisely, under the assumption of letter-based word recognition, exposure to printed words should affect the processing of individual letters as a function of their frequency of occurrence in words. What is the evidence for this?

While word frequency effects are one of the most robust and widely studied phenomena in cognitive psychology, relatively few studies have investigated effects of letter frequency. Furthermore, the majority of these studies failed to find a letter frequency effect. Most of these early studies used the same-different matching task (see Appelman & Mayzner, 1981, for review). In this task participants see two letters (simultaneously or in sequence) and respond to indicate if they are the same or different. The absence of a letter frequency effect in these studies could therefore be due to participants matching letters by using shape information (have the same visual form) rather than letter identities (are the same letter), as pointed out by Posner and Mitchell (1967) and Appelman and Mayzner (1981).

Appelman and Mayzner (1981) concluded from their re-analysis of 58 studies using percent errors as the dependent variable, that letter frequency does not affect the perceptual processing of letters. For studies using response times (RTs) as dependent variable, they found only 3 experiments out of 6 reported in two studies (Cosky, 1976; Podgorny & Gardner, 1979) where letter frequency correlated significantly with RTs. One of these experiments used the same-different matching task (Podgorny & Gardner, 1979), one used a letter naming task (Cosky, 1976), and one used a letter/non-letter discrimination task (Cosky, 1976). We have already mentioned the problems associated with interpreting results obtained with the same-different matching task. The naming task is also problematical in that much of the variance in naming latencies is driven by the nature of the initial phoneme (Rastle

* Corresponding author at: Institut de Psychologie, 71 Avenue Edouard Vaillant, F-92774 Boulogne-Billancourt, France. Tel.: +33 1 5520 5864.
E-mail address: boris.new@parisdescartes.fr (B. New).

[1] Varying resting level activation is just one way of implementing the general principle of statistical learning in connectionist models.

0001-6918/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2011.07.001

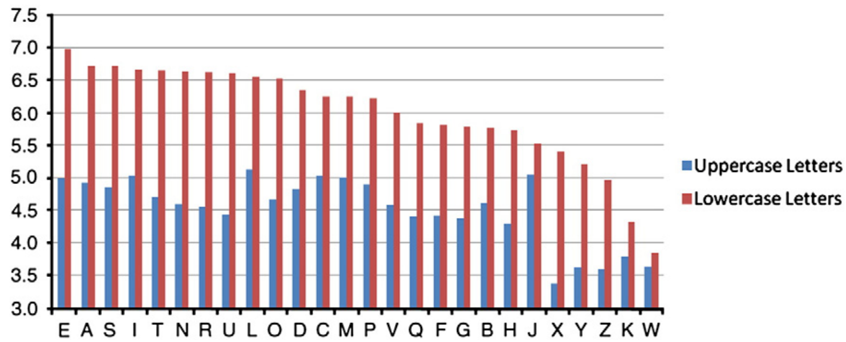


Fig. 1. Log frequency of occurrence in French for uppercase and lowercase formats of the 26 letters of the Roman alphabet (ranked by lowercase frequency) calculated from the Frantext corpus used for Lexique 3.55 (New, Pallier, Ferrand, & Matos, 2001; New, Pallier, Brysbaert, & Ferrand, 2004).

& Davis, 2002). Arguably the most convincing demonstration of a letter frequency effect was that seen in letter/non-letter discrimination times. However, there are a number of reasons to be cautious about this finding. First, it is based on a re-analysis of Cosky's (1976) data by Appelman and Mayzner (1981), where they averaged RTs across two different conditions tested by Cosky: one where the non-letters were non-alphabetic characters (e.g., # , *), and one where the non-letters were upside-down rotated or mirror image letters. Furthermore, only uppercase letters were tested.

Further evidence for a letter frequency effect was recently reported by Pitchford, Ledgeway, and Masterson (2008) in a study using the letter search task. In this study, participants had to detect a pre-determined target letter among a string of 5 letters. Targets were present in the string on 50% of trials, and could appear at one of the 5 possible positions in the string. Pitchford et al. observed the typical M-shaped serial position function of target detection latencies in their experiments. Most important, for the purposes of the present study, is that RTs to individual letters averaged over the 5 possible positions in the string were found to correlate significantly with letter frequency. However, when analyzed separately for each position in the string, only RTs to letters tested at the initial and final positions of strings correlated significantly with letter frequency. Further work using the letter search task has revealed that the letter frequency effect for the final position in the string, reported by Pitchford et al. (2008) for skilled readers of English, is not seen with dyslexic readers of English of the same educational level (Pitchford, Ledgeway, & Masterson, 2009), nor with skilled readers of Greek (Ktori & Pitchford, 2009). Letter frequency did, however, correlate with RTs to detect targets at initial position of letter strings in these studies.

Given the theoretical importance of letter frequency effects, and the relative scarceness of research on this topic up to now, the present study provides a further investigation of this key issue (see Rey, Dufau, Massol, & Grainger, 2009, for a recent study of letter identification, and Grainger, Rey, & Dufau, 2008, for a review of the literature). Within this general perspective, the present study has one general goal and two more specific goals. Our general goal is to replicate the letter frequency effect revealed by Appelman and Mayzner in their re-analysis of Cosky's (1976) data. This is clearly an important goal given that Appelman and Mayzner actually concluded that letter frequency does not affect letter perception. Furthermore, our study examines the presence of letter frequency effects in another language (French instead of English) and with other types of non-letters. We first replicate the letter frequency effect with uppercase letters (as in Cosky's study), before testing its generalizability to lowercase letters. Indeed, there are reasons to believe that the effect might be more difficult to obtain with lowercase letters. First, lowercase letters are almost never presented in isolation. Second, lowercase letters occur much more frequently than uppercase letters, and this could lead to a ceiling effect. Finally, apart from demonstrating a letter frequency effect in the alphabetic decision task with uppercase and lowercase letters, in the present study we seek to: 1) Compare the correlations obtained with several different frequency measures – type frequency, token frequency, frequency rank; and 2) Investigate the influence of the type of corpus on which the letter frequency counts are derived (printed texts vs. film subtitles).

Concerning the different frequency measures, the type frequency for the letter "a" for instance is the number of words that contain the letter "a". Contrary to token frequency counts, type frequency is not weighted by word frequency. Connectionist models of letter and word processing (such as the interactive-activation model) predict that letter perception should be maximally sensitive to token frequency, since it is the number of times a letter is seen that is important, independently of the surrounding context. Moreover, general accounts of frequency effects as reflecting frequency-ordered search mechanisms predict that it is frequency rank that should best predict performance (for an account of word frequency effects as rank effects, see Murray & Forster, 2004). Finally, accounts of frequency effects as reflecting contextual diversity, that is more frequently occurring items also occur in a greater number of different contexts (for an account of word frequency effects as effects of contextual diversity, see Adelman, Brown, & Quesada, 2006), predict that it is type frequency that should best predict performance.

Concerning possible effects of the nature of the corpus used to calculate letter frequencies, we will compare letter frequencies calculated from a corpus of written materials (essentially novels) vs. those calculated from a corpus of film subtitles (used to estimate spoken language frequencies). Interestingly, in prior work it was found that subtitle frequency is a better predictor of lexical decision times than book frequency (Brysbaert & New, 2009; Ferrand et al., 2010; Keuleers, Brysbaert, & New, 2010; New, Brysbaert, Veronis, & Pallier, 2007). Although the basis of this difference remains to be clarified, here we predict that contrary to lexical decision, alphabetic decision latencies should be more sensitive to printed word frequency compared with spoken word frequency.

Experiment 1A and 1B first test for frequency effects with isolated letter stimuli in a speeded binary decision task involving letter vs. pseudo-letter classification. We chose this task as the equivalent of the lexical decision task, since the lexical decision task is probably the most widely used task in the study of visual word recognition. Furthermore, there is evidence that the alphabetic decision task is sensitive to basic processes in letter identification (Jacobs & Grainger, 1991; Jacobs, Grainger, & Ferrand, 1995; Peressotti & Grainger, 1995; Ziegler, Ferrand, Jacobs, Rey, & Grainger, 2000).

1. Experiment 1: isolated letters

1.1. Participants

Sixteen students from the Université de Provence took part in each experiment for course credit. They were all native French speakers and had normal or corrected-to-normal vision.

1.2. Stimuli and design

We presented 18 different consonants (B, C, D, F, G, H, K, L, M, N, P, Q, R, S, T, V, W, Z) in the center of the screen, and 18 pseudo-letters designed using Font Creator 4.0 software (see Appendix A for a sample of uppercase pseudo-letters). The pseudo-letters stimuli were distortions of their corresponding letters of the Roman alphabet. Pseudo-letters were not rotations or mirror images of real letters.

The same letters and pseudo-letters were presented 6 times each giving 6 observations per stimulus per participant. The stimuli in Experiment 1B were the same as that in Experiment 1A except that we used lowercase letters instead of uppercase letters (see Appendix B for a sample of lowercase pseudo-letters). A different set of pseudo-letters was also derived from lowercase letters. The token and type frequencies were calculated separately for these lowercase and uppercase letters in the FRANTEXT corpus composed of 14.7 million words (the corpus used for computing word frequencies in Lexique 3). In this way, for instance, uppercase letters in proper names are counted in our uppercase letter frequency measure but not in our lowercase letter frequency measure. Therefore separate case-specific frequency counts were used to predict performance on lowercase and uppercase letters in Experiments 1A and 1B.

1.3. Procedure

Participants were tested individually in a quiet room. They were asked to indicate as quickly and accurately as possible whether the presented stimulus was an existing letter or not. They did so by pressing "l" or "q" on the keyboard. Each trial began with a 200 ms fixation cross (a plus sign in the center of the screen), followed by the stimulus, which remained visible until the participant responded (with a time-out of 4 s). Between trials, there was a 1300 ms black screen interval. Trials were randomized anew for each participant and presented using E-Prime 1.1 (Psychology Software Tools). The experimental trials were preceded by twelve practice trials using 6 letters (a, e, i, o ,u y) and 6 pseudo-letters not tested in the main experiment, each presented twice. Only consonants were tested in the main experiment because it has been shown that vowels could have a special status during reading (New, Araujo, & Nazzi, 2008). Stimuli were displayed using "Courier New" font. Participants could take a short break between the two blocks composed of 108 trials each. The experiment lasted approximately 15 min.

1.3.1. Results & discussion

Only RTs of correct responses were included in the RT analyses. In addition, RTs greater than two standard deviations above or below the participant's mean were discarded as outliers (4.1% of the data for experiment 1A and 1B). Adjusted R-squared values computed from the linear regression of our different letter frequency measures on RT are presented in Tables 1 and 2. All correlations were negative.

1.3.1.1. Experiment 1A: uppercase letters. The key result with uppercase letters is that token frequency calculated across all positions ($F(1, 16)=81.4$) or just initial position ($F(1, 16)=75.6$; $p<0.001$) explained the largest amount of variance (more than 80%). Alphabetic decision latencies decreased as letter frequency increased. Furthermore, concerning overall letter frequency and initial letter frequency, token frequency explained more variance than either type frequency or frequency rank.

Table 1
Effects (R-squared values) of log-transformed token and type frequency, and rank token frequency on alphabetic decision latencies in Experiments 1A (uppercase letter targets) and 1B (lowercase letter targets). Frequency position refers to the within-word positions used to calculate letter frequency.

Frequency position	Uppercase letters			Lowercase letters		
	Token	Type	Rank	Token	Type	Rank
All Positions	0.836 ***	0.759 ***	0.675 ***	0.375 **	0.344 *	0.226 *
Initial	0.825 ***	0.695 ***	0.652 ***	0.44 **	0.372 **	0.387 **
Interior	0.748 **	0.771 ***	0.667 ***	0.324 *	0.323 *	0.248 *
Final	0.396 **	0.436 **	0.405 **	0.098	0.021	0.094

Note. * $p<.05$. ** $p<.01$. *** $p<.001$.

Table 2
Effects (R-squared values) of log-transformed token frequency based on the book or the subtitle corpus, computed on all positions or initial position only.

Letter position frequency	Uppercase letters		Letter position frequency	Lowercase letters	
	Books	Subtitles		Books	Subtitles
All positions	0.836 ***	0.674 ***	All positions	0.375 **	0.355 **
Initial	0.825 ***	0.664 ***	Initial	0.44 **	0.407 **

Note. * $p<.05$. ** $p<.01$. *** $p<.001$.

1.3.1.2. Experiment 1B: lowercase letters. For lowercase letters, a main effect of token frequency was also obtained, with alphabetic decision latencies being faster for more frequently occurring letters. This effect was also significant for overall ($F(1, 16)=9.61$; $p<0.01$) and initial letter frequency ($F(1, 16)=12.6$; $p<0.01$). As for uppercase letters, we also observed that initial and overall letter frequency were the best predictors. We also found that token frequency was a better predictor than either type frequency or frequency rank.

Experiment 1 revealed a robust influence of letter frequency on performance in the alphabetic decision task with isolated letters and pseudo-letters. We found a significant letter frequency effect for both lowercase and uppercase letters. Furthermore, four main patterns emerged from this first experiment: 1) correlations were higher with uppercase letters than lowercase letters; 2) initial letter frequency generated greater correlations than either final letter frequency or interior letter frequency; 3) token frequency was a better predictor than type frequency, which in turn was a better predictor than frequency rank; 4) letter frequencies based on the book corpus correlated better with reaction times than letter frequencies based on the subtitle corpus.

2. Experiment 2: letters embedded in strings

Experiment 2 provides a replication of the letter frequency effect, but this time with letters presented in strings. More specifically Experiment 2 examines whether the frequency effects obtained with isolated letters can be generalized to the situation where letter targets are embedded in character strings. There is prior evidence that processing letters in strings might involve different mechanisms compared with isolated letter perception. For example, Grainger and Jacobs (1991) used the alphabetic decision task and the masked priming paradigm with word primes and letter targets (e.g., table – T). They failed to find a priming effect when the target letter was an isolated letter ("T"), but found a robust priming effect when the target letter was embedded in a character string ("T####"). Experiments 2A and 2B therefore test whether effects of letter frequency in the alphabetic decision task are also obtained when letter targets are embedded in strings, and examine whether letter frequency effects are sensitive to the target letter's location in the string.

2.1. Method

2.1.1. Participants

Sixteen students from the Université René Descartes, Paris V, took part in each experiment in return for course credit. They were all native French speakers and had normal or corrected-to-normal vision.

2.1.2. Stimuli and design

All stimuli consisted of horizontal arrays of five characters made of a given target letter and four Xs. The conditions were the same as in Experiment 1A except that stimuli could appear in 3 different positions (beginning ("CXXXX"), middle ("XXCXX") or end ("XXXXC")). The same letters and pseudo-letters were presented in 2 different blocks. Thus RTs for each consonant at each target position were based on 2 observations per participant. The letter and pseudo-letter stimuli were the same as in Experiment 1.

2.1.3. Procedure

The procedure in Experiment 2 was very similar to Experiment 1. Each trial began with a 200 ms fixation cross (a plus sign in the center of the screen), followed by the stimulus, which remained visible until the participant responded (with a time-out of 4 s). Between trials, there was a 1300 ms black screen interval. Trials were randomized anew for each participant and presented using E-Prime 1.1.

2.2. Results & discussion

Only RTs of correct responses were included in the RT analyses. In addition, RTs greater than two standard deviations above or below the participant's mean were discarded as outliers (4.8% of the data). We also removed the results for the letter "K" as the error rate was more than 30% for that letter (likely due to confusion with the letter X used as a filler). Adjusted R-squared values computed from the linear regression of our different letter frequency measures on RT are presented in Table 3. All correlations were negative.

2.2.1. Experiment 2A: uppercase letters

The results show that initial token frequency correlated significantly with RTs when the target was presented at the beginning of the string ($F(1, 15) = 5.64$; $p < 0.05$). Token frequency computed across all positions also correlated significantly with RTs when letters were presented at the beginning of strings ($F(1, 15) = 5.38$; $p < 0.05$). Final and interior token frequency did not correlate significantly with RTs at any position. Comparing the different frequency measures calculated for initial position in words, token frequency explained more variance than type frequency, but frequency rank explained more variance than token frequency. Finally, frequency calculated from the book corpus generated stronger correlations than the subtitle frequency counts.

Table 3
Effects (R-squared values) of log-transformed token and type frequency, and token frequency rank on alphabetic decision latencies with target letters embedded in a string of Xs at either beginning, middle, or end position in Experiments 2A (uppercase letters) and 2B (lowercase letters). Frequency position refers to the within-word positions used to calculate letter frequency. Type frequency, frequency rank, and subtitle token frequency were calculated for initial position only.

Uppercase letters			
Frequency position	Target letter position		
	Initial ("CXXXX")	Medial ("XXCXX")	Final ("XXXXC")
All positions	0.264 *	0.096	0.156
Initial	0.266 *	0.096	0.155
Interior	0.188	0.059	0.098
Final	0.001	0.043	0.035
Type	0.256 *	0.094	0.111
Rank	0.29 *	0.041	0.139
Subtitles	0.13	0.064	0.075
Lowercase letters			
Frequency position	Target letter position		
	Initial ("cxxxx")	Medial ("xxcxx")	Final ("xxxxc")
All positions	0.171 .	0.169 .	0.211 .
Initial	0.261 *	0.267 *	0.32 *
Interior	0.129	0.122	0.136
Final	0.026	0.007	0.055
Type	0.178 .	0.165 .	0.178 .
Rank	0.129	0.111	0.259 *
Subtitles	0.219 *	0.221 *	0.30 *

Note. * $p<.05$. ** $p<.01$ *** $p<.001$. $p<.1$.

2.2.2. Experiment 2B: lowercase letters

The results show that initial token frequency correlated significantly with RTs when the target was presented at the beginning of the string ($F(1, 16) = 5.65$; $p < 0.05$), at the middle ($F(1, 16) = 5.84$; $p < 0.05$), or at the end position ($F(1, 16) = 7.52$; $p < 0.05$). Token frequency computed across all positions did not reach significance whatever the target letter position (beginning ($F(1, 16) = 3.30$; $p = 0.088$), middle ($F(1, 16) = 3.24$; $p = 0.09$) or end ($F(1, 16) = 4.27$; $p = 0.055$). Final and interior token frequency did not correlate significantly with RTs in any position. Finally, neither initial type frequency nor initial frequency rank explained more variance than initial token frequency. As with uppercase letters, frequency calculated from the book corpus generated stronger correlations than the subtitle frequency counts.

The results of Experiment 2 show that letter frequency effects in the alphabetic decision task can also be found when letter targets are embedded in strings of Xs. There were five main patterns in the results of this experiment: 1) correlations were higher with lowercase letters than uppercase letters, except for targets appearing at the beginning of strings; 2) initial letter frequency generated greater correlations than either final letter frequency or interior letter frequency; 3) token frequency was almost always a better predictor than either type frequency or frequency rank; 4) correlation with uppercase letters were only significant for targets at the beginning of strings, whereas lowercase letters showed significant correlations at all positions; 5) correlations were greater for frequency counts based on the book corpus compared with subtitle frequency counts.

3. Combined analyses of Experiments 1 and 2

In order to provide statistical tests of observed differences in the strength of the correlations with RT and our different letter frequency measures, we performed a combined analysis of the results of Experiments 1 and 2 after transforming the RT and the log of the frequencies into z scores. Concerning the different positions tested in Experiment 2 (initial, medial, final), with lowercase letters we used the RT averaged across the 3 positions, since our results were similar for the different positions. For uppercase letters, we only considered letters presented at initial position, as this was the only position to generate significant frequency effects. In the combined analysis, when our dependent variable was the same (typically when comparing two different frequency measures), we used Clarke's test (Clarke, 2007) since the log likelihoods for our regression models were not normally distributed. Clarke's (2007) distribution-free test applies a modified paired sign test for the differences in the individual log-likelihoods from two non-nested regression models. As Clarke's comparison requires the same dependent variable, when our dependent variable was not the same (such as when comparing lowercase and uppercase letters, and comparing isolated letters vs. letters presented in context), we performed regressions on individual subjects' z scores and then extracted the beta weights from these analyses and performed t-tests on the beta weights (Lorch & Myers, 1990).

First, we compared letter frequency measures based on initial letter frequency, vs. interior letter frequency and final letter frequency. These analyses were performed using token frequency given the superior correlations obtained with token frequency counts. The results showed that the initial letter frequency model generated significantly higher R-squared values than both interior letter frequency ($p < 0.001$) and final letter frequency ($p < 0.001$). Next, we compared the models obtained with token frequency, type frequency, and frequency rank. The results showed that the model with token frequency was significantly preferred compared to the models with either type frequency ($p < 0.05$) or rank frequency ($p < 0.001$). We also found that the models obtained with token frequency counts derived from a book corpus were significantly preferred over the models obtained with token frequency counts derived from a corpus of film subtitles ($p<0.001$). This result was replicated when we limited the analysis to only lowercase letters ($p<0.05$) or only uppercase letters ($p<0.001$). Next, we compared our regression models obtained with isolated letters in Experiment 1 to those obtained with letters embedded in strings of Xs in Experiment 2. The results revealed that overall the standardized beta weights were higher for isolated letters than for letters embedded in strings ($t(31)=-2.3$; $p<0.05$). Finally, we compared the beta weights obtained for uppercase letters and lowercase letters in Experiments 1 and 2. This analysis revealed that the beta weights obtained with uppercase letters did not differ significantly from those obtained with lowercase letters, even in the isolated presentation conditions of Experiment 1, where the biggest differences in the size of correlations was observed ($t(15)=-1.08$; $p=0.15$).

3.1. General discussion

The present experiments provide clear evidence for a letter frequency effect in the alphabetic decision task. In Experiment 1A, letter frequency effects were found for uppercase isolated letter targets. In Experiment 1B, letter frequency effects were generalized to lowercase letters. Experiments 2A and 2B replicated these findings and extended them to the situation where letters were embedded in strings of Xs and could appear at the beginning, middle, or final position of a 5-character string. Finding a robust effect of letter frequency has important theoretical consequences that will be developed below.

Apart from clearly establishing effects of letter frequency for isolated uppercase and lowercase letters and letters embedded in strings, the main findings of the present study can be summarized as follows:

1) Initial letter frequency generated stronger correlations than either final letter frequency or interior letter frequency for isolated letters and letters in strings.
2) Correlations were higher for frequency counts derived from a book corpus compared with the frequency counts derived from a corpus of film subtitles.
3) Token frequency was a better predictor than both type frequency and frequency rank.
4) Standardized beta weights were significantly higher for isolated letters than for letters embedded in strings.
5) For letters embedded in strings, the correlations with uppercase letters were only significant for targets at initial position, whereas lowercase letters showed significant correlations at all positions.
6) For isolated letters, correlations were greater for uppercase letters than lowercase letters, but the standardized beta weights were found not to be significantly different.

3.1.1. Effects of letter frequency

Robust effects of letter frequency were found in the present study. First, we replicated the letter frequency effect found with uppercase letters in isolated presentation in Appelman and Mayzner's (1981) reanalysis of Cosky's data. Second, we found that this letter frequency effect generalized to lowercase letters. Third, we found a letter frequency effect for letters presented in strings of Xs. These letter frequency effects were found to vary as a function of the way in which letter frequency was measured.

The finding that token frequency is a better predictor than type frequency provides support for the idea that every time a word is read its constituent letters are processed to a certain extent, and that this processing modifies the future accessibility of the letter representations that are involved (via modifications of connection strengths or resting-level activations, for example). It should be noted, however, that this does not necessarily imply that words composed of more frequent letters will be easier to identify than words composed of less frequent letters. The problem is that more frequent letters are less informative with respect to word identity than less frequent letters, hence explaining why certain orthographic priming effects are smaller when primes and targets share high-frequency letters compared with low-frequency letters (e.g., Lupker, Perea, & Davis, 2008).

Accounts of frequency effects as reflecting contextual diversity predicted superior correlations with type frequency. According to this account, more frequently occurring items also occur in a greater number of different contexts, and it is this increased contextual diversity that aids performance (Adelman et al., 2006). Therefore, it should be the number of different words in which a given letter appears (type frequency) that best predicts performance to letter targets. This was not the case in the present study.

Furthermore, general accounts of frequency effects as reflecting frequency-ordered search mechanisms, predict that it is frequency rank that should best predict performance (e.g., Murray & Forster's, 2004, account of word frequency effects). Applied to the case of letter identification, the general principle is that perceptual information (i.e., visual features) is matched to letter representations in long-term memory sequentially, one letter at a time, in the order of the frequency of occurrence of letters. Letter identification times would therefore be a function of the position of a given letter in this frequency ranking. The lower correlations with frequency rank compared with token frequency provide evidence against this general approach to explaining frequency effects.

We also investigated a possible influence of the type of corpus on which our letter frequency counts were derived. More precisely, we compared letter frequencies calculated from a corpus of written materials (essentially novels) vs. those calculated from a corpus of film subtitles (used to estimate spoken language frequencies). Interestingly, in prior work it was found that subtitle frequency is a better predictor of lexical decision times than book frequency (Brysbaert & New, 2009; Ferrand et al., 2010; Keuleers et al., 2010; New et al., 2007). Exactly the opposite effect was found for letter frequencies and alphabetic decision latencies in the present study. Book frequency was systematically a better predictor than subtitle frequency for lowercase and uppercase letters. For instance in Experiments 1A and 1B subtitle initial token frequency explained respectively 64% and 37% of variance, while book initial token frequency explained 81% and 40%. The same results were found for experiments 2A and 2B (7.2% and 17% for subtitle frequency compared to 21.7% and 21.9% for book frequency). This could simply be due to the fact that spoken language processing does not involve activation of letter-level representations to the same extent as the processing of written language.

3.1.2. Effects of letter format and letter position

Letter frequency was found to explain a large amount of variance in RT to isolated uppercase letters in Experiment 1, and these correlations were larger than those found for isolated lowercase letters, although the beta weights of the individual regressions did not differ significantly. Furthermore, the numerical advantage for uppercase letters disappeared in Experiment 2 when letters were embedded in strings of Xs. In these conditions the correlations were overall weaker, and when letters were presented in the centre or at the end of strings, then correlations were numerically greater for lowercase than uppercase letters. Future research should aim to clarify to what extent correlations with letter frequency differ for uppercase and lowercase formats. It is possible that greater exposure to isolated letters in uppercase than lowercase format might indeed lead to greater correlations with letter frequency.

Identification of letter targets embedded in strings of Xs was also significantly influenced by letter frequency. These letter frequency effects were found at all target positions (initial, medial, final) with lowercase letters, and for only the first position in the string for uppercase targets. It is possible that the position-dependent nature of the frequency effect for uppercase targets is due to the fact that uppercase letters are generally seen at the initial position of words. If this were indeed the case, then one might expect to see a stronger correlation with letter frequency when the frequency count is limited to the number of occurrences of letters as the first letter of the first word of a sentence. However, additional analyses showed that this measure of letter frequency actually explains less variance (11% instead of 22%). This suggests that the restricted nature of the correlations seen with uppercase letters in context, is more related to general mechanisms that reflect the fact that uppercase letters generally appear at the first position of words, as opposed to specific exposure to letters in uppercase format. One such general mechanism would involve preparation for processing uppercase letters at the initial position in a string of letters.

Separate letter frequency values were calculated for letters appearing at the beginning of words (initial position), the end of words (final position), and for interior letters. The correlations with these different position-specific frequency measures revealed that frequency of occurrence as the first letter of words was the most effective variable in terms of modulating letter identification times, both for isolated letters and letters embedded in strings. Frequency of letters at the last position in words or as interior letters, showed much lower correlations with alphabetic decision latencies. Our explanation of this effect of type of letter frequency measure is based on the well-established fact that letters at the initial position in a word are generally the most visible (e.g., Stevens & Grainger, 2003). If one assumes that effects of frequency of exposure are the result of modifications of connection strengths, then more visible letters will therefore generate greater changes in the strength of feature-letter connections, leading to greater effects of letter frequency. Finally, another possibility is that the difference between upper and lowercase letters in Experiment 2 reflects the fact that uppercase letters were harder to find in a string of uppercase Xs than were lowercase letters. As letters in the initial position are more visible, this could explain why correlations with letter frequency were observed for uppercase letters in this position.

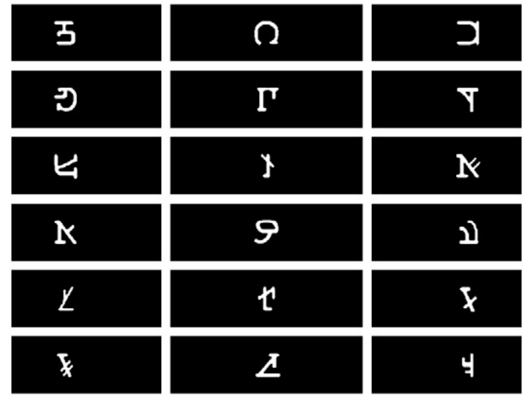
3.1.3. Conclusions

Summing up, we found that letter frequency correlated significantly with RTs both to uppercase and lowercase letter targets in an alphabetic decision task. We found that token frequency was a better predictor than either type frequency or frequency rank, thus lending support to letter-based models of word recognition such as the interactive activation model, where letter units are systematically activated when a word is read. We found that frequency counts derived from a book corpus generated stronger correlations than the frequency counts from a corpus of film subtitles, suggesting that exposure to print is the critical factor underlying letter frequency effects. We found that frequency of occurrence of letters at the beginning of words was a better predictor than frequency of interior or final letters. This was interpreted as reflecting the improved processing of initial letters in strings, which in turn would lead to greater sensitivity to frequency of occurrence. Finally, correlations with frequency varied as a function of letter format (uppercase vs. lowercase) and position of the target letter in a string of Xs, thus providing evidence for the existence of position-specific and case-specific letter processing mechanisms.

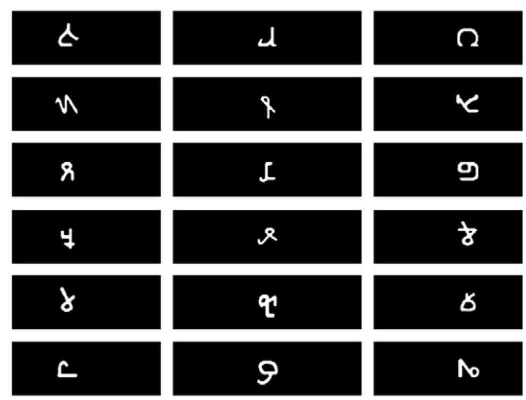
Acknowledgements

We would like to thank Magali Boibieux and Yousri Marzouki for their contributions. This research was supported by ERC grant 230313.

Appendix A. Uppercase pseudo-letters used in Experiment 1A and 2A



Appendix B. Lowercase pseudo-letters used in Experiment 1B and 2B



Appendix C. Letter targets and corresponding mean RTs in Experiments 1A & 1B

Experiment 1A		Experiment 1B	
Letter	RT	Letter	RT
B	413	b	432
C	412	c	439
D	413	d	428
F	419	f	469
G	419	g	444
H	431	h	455
K	436	k	460
L	412	l	435
M	406	m	461
N	420	n	440
P	414	p	425
Q	429	q	453
R	412	r	458
S	419	s	450
T	416	t	441
V	422	v	449
W	445	w	479
Z	437	z	463

Raw results are available as supplementary data.

Appendix D. Letter targets and corresponding mean RTs for each position in Experiments 2A & 2B

Experiment 2A				Experiment 2B			
Letter	Initial	Medial	Final	Letter	Initial	Medial	Final
B	504	498	465	b	497	475	478
C	507	498	463	c	508	483	473
D	507	500	480	d	494	470	456
F	521	475	519	f	499	476	469
G	482	499	468	g	498	484	480
H	529	524	523	h	510	474	484
L	479	493	467	k	515	526	522
M	518	528	540	l	503	469	456
N	541	521	520	m	519	522	537
P	493	492	486	n	516	500	482
Q	524	499	498	p	511	484	464
R	540	506	501	q	503	482	485
S	492	481	477	r	547	515	535
T	501	520	486	s	525	511	465
V	540	489	491	t	488	484	478
W	524	510	562	v	528	492	474
Z	539	526	478	w	562	529	523
				z	527	493	501

Appendix E. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.actpsy.2011.07.001.

References

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, *17*, 814–823.

Appelman, I. B., & Mayzner, M. S. (1981). The letter-frequency effect and the generality of familiarity effects on perception. *Perception & Psychophysics*, *30*, 436–446.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990.

Clarke, K. A. (2007). A simple distribution-free test for nonnested hypotheses. *Political Analysis*, *15*, 347–363.

Cosky, M. J. (1976). The role of letter recognition in word recognition. *Memory & Cognition*, *4*, 207–214.

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., et al. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, *42*, 488–496.

Grainger, J. (2008). Cracking the orthographic code: An introduction. *Language and Cognitive Processes*, *23*, 1–35.

Grainger, J., & Jacobs, A. M. (1991). Masked constituent letter priming in an alphabetic decision task. *European Journal of Cognitive Psychology*, *3*, 413–434.

Grainger, J., Rey, A., & Dufau, S. (2008). Letter perception: from pixels to pandemonium! *Trends in Cognitive Sciences*, *12*, 381–387.

Jacobs, A. M., & Grainger, J. (1991). Automatic letter priming in an alphabetic decision task. *Perception & Psychophysics*, *49*, 43–52.

Jacobs, A. M., Grainger, J., & Ferrand, L. (1995). The incremental priming technique: A method for determining within-condition priming effects. *Perception & Psychophysics*, *57*, 1101–1110.

Keuleers, E., Brysbaert, M., & New, B. (2010). Subtlex-nl: A new frequency measure for Dutch words based on film subtitles. *Behavior Research Methods*, *42*, 643–650.

Ktori, M., & Pitchford, N. J. (2009). Development of letter position processing: Effects of age and orthographic transparency. *Journal of Research in Reading*, *32*, 180–198.

Lorch, R. F., Jr., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research: A comparison of three different methods. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *16*, 149–157.

Lupker, S. J., Perea, M., & Davis, C. J. (2008). Transposed-letter effects: Consonants, vowels, and letter frequency. *Language and Cognitive Processes*, *23*, 93–116.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375–407.

Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, *111*, 721–756.

New, B., Araujo, V., & Nazzi, T. (2008). Differential processing of consonants and vowels in lexical access through reading. *Psychological Science*, *19*, 1223–1227.

New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied PsychoLinguistics*, *28*, 661–677.

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A New French Lexical Database. *Behavior Research Methods, Instruments, & Computers*, *36*, 516–524.

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE. *L'Année Psychologique*, *101*, 447–462. http://www.lexique.org

Peressotti, F., & Grainger, J. (1995). Letter position coding in random consonant arrays. *Perception & Psychophysics*, *57*, 875–890.

Pitchford, N. J., Ledgeway, T., & Masterson, J. (2008). Effect of orthographic processes in letter position encoding. *Journal of Research in Reading, Special Issue: Orthographic Processes in Reading*, *31*, 97–116.

Pitchford, N. J., Ledgeway, T., & Masterson, J. (2009). Reduced orthographic learning in dyslexic adult readers: Evidence from patterns of letter search. *Quarterly Journal of Experimental Psychology*, *62*, 99–113.

Podgorny, P., & Gardner, W. R. (1979). Reaction time as a measure of inter-object visual similarity: Letters of the alphabet. *Perception & Psychophysics*, *26*, 37–52.

Posner, M. J., & Mitchell, R. F. (1967). Chronometric analysis of classification. *Psychological Review*, *74*, 392–409.

Rastle, K., & Davis, M. H. (2002). On the complexities of measuring naming. *Journal of Experimental Psychology. Human Perception and Performance*, *28*, 307–314.

Rey, A., Dufau, S., Massol, S., & Grainger, J. (2009). Testing computational models of letter perception with item-level ERPs. *Cognitive Neurospsychology*, *26*, 7–22.

Stevens, M., & Grainger, J. (2003). Letter visibility and the viewing position effect in visual word recognition. *Perception & Psychophysics*, *65*, 133–151.

Ziegler, J., Ferrand, L., Jacobs, A. M., Rey, A., & Grainger, J. (2000). Visual and phonological codes in letter and word recognition: Evidence from incremental priming. *Quarterly Journal of Experimental Psychology*, *53A*, 671–692.